Computer Architecture 2024/25

Homework TPC1

This homework consists of a programming exercise to be carried out in a group of no more than two students. You can clarify general doubts with colleagues, but the solution and writing of the code should be strictly carried out by the members of the group. Cases of plagiarism will be handled in accordance with the regulations in force. It is possible to use tools such as CoPilot or ChatGPTand it is not mandatory to report the use; don't forget that <u>the mark in this exercise will be validated by dedicated questions in 1st test</u>.

Deadline for delivery: April, 1st (Tuesday) at 5:00PM.

Delivery will be done through Mooshak. Details about the delivery will be sent by e-mail sent through CLIP.

Exercise

A file is a sequence of bytes stored in a disk, being designated by a name which is a character string. Regarding its contents, a file can be classified in two distinct categories:

- **Text**: file contains only bytes that represent something that can be displayed in a terminal: in this exercise, those bytes are:
    - Bytes with values 9 (tab '\t'), 10 (carriage return, '\r') and 13 (line feed '\n')
    - Bytes with values between 32 and 64 corresponding to characters like space, '*', '!', and others.
    - Bytes with values between 65 and 90: the capital letters 'A' to 'Z'
    - Bytes with values between 91 and 96: more characters, namely '\' , '^', '[' , ']'
    - Bytes with values between 97 and 122: small case letters 'a' to 'z'
    - Bytes with values between 123 and 126: more characters like '{', '}', '|', '^', '~'
    - Bytes with values between160 and 255: accentuated letters and other symbols
- **Binary**: if the file contains a byte not included in the above list.
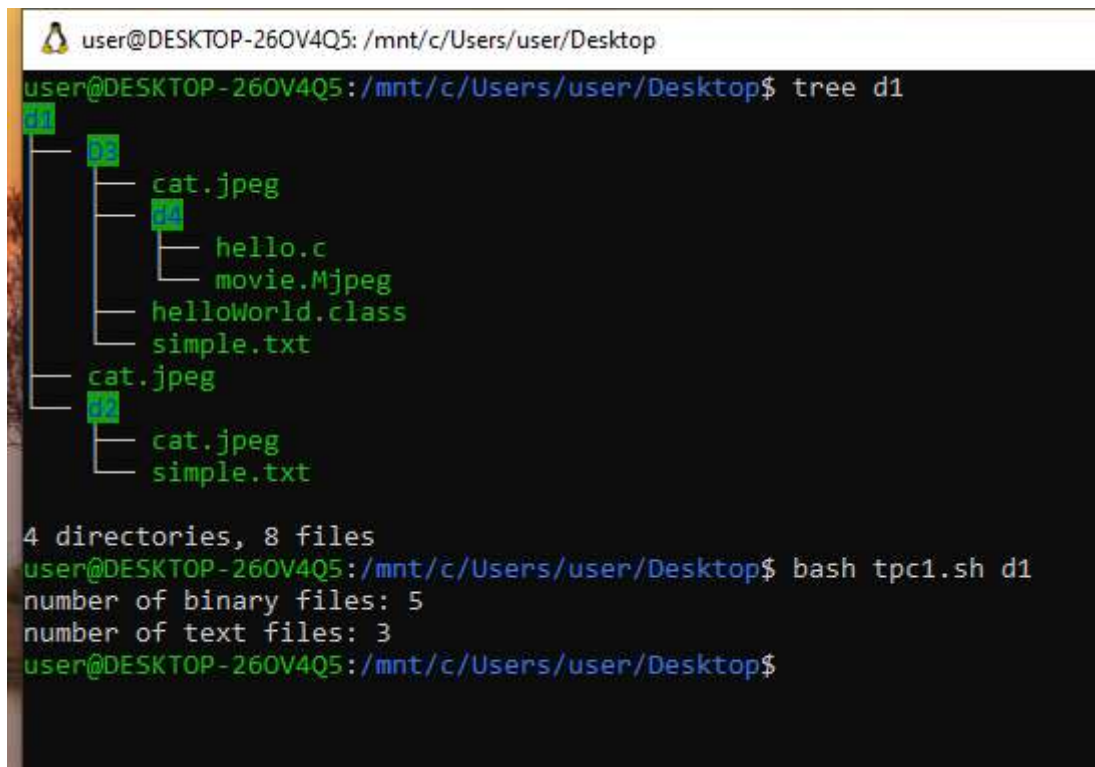
The practical work consists of creating two programs:

- a C program called i*sText* that accepts a file name as an input argument, and returns 1 if the file is a text file and 0 if it is a binary file, according to the rules indicated above.

- a bash script called *tpc1.sh* that finds all the files in a given sub-tree of the file systems and invokes the *isText* program for each of the files found. According to the value returned by *isText,* the script updates two counters: one for text files and another for binary files. At the end, the script prints the total number of files of each type.

File *tpc1.zip*, available from CLIP, a set of folders and files, where:
- *hello.c* and *simple.txt* are text files.
- *cat.jpeg, helloWorld.class,* and *movie.Mpeg* are binary files.

The following figure shows the contents of a subtree of the file system, obtained through the command *tree*[1]; the result of the execution of script *tpc1.sh* is shown after.


```
△ user@DESKTOP-26OV4Q5: /mnt/c/Users/user/Desktop
user@DESKTOP-26OV4Q5:/mnt/c/Users/user/Desktop$ tree d1
d1
├── D3
│   ├── cat.jpeg
│   ├── d4
│   │   ├── hello.c
│   │   └── movie.Mjpeg
│   ├── helloWorld.class
│   └── simple.txt
├── cat.jpeg
└── d2
    ├── cat.jpeg
    └── simple.txt

4 directories, 8 files
user@DESKTOP-26OV4Q5:/mnt/c/Users/user/Desktop$ bash tpc1.sh d1
number of binary files: 5
number of text files: 3
user@DESKTOP-26OV4Q5:/mnt/c/Users/user/Desktop$
```

---

[1] In Ubuntu you can install the program *tree,* using the command *sudo apt-get install tree* .